

**Exporting Statistical Expertise
to Developing Countries: Discussion**

by

Walter T. Federer

BU-965-M

April, 1988

ABSTRACT

Three papers by L. Kish, J. C. Arnold, and R. McCracken were presented on a Session at the August 1986 Joint Statistics Meetings. Some comments on the topic as well as comments on each of the three papers are presented. Deficiencies in present statistics course offerings are noted.

* In the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, NY 14853. This Discussion with the three papers is published in a special series on public health education in developing countries by the John J. Sparkman International Center for Public Health Education.

Discussion

Walter T. Federer, Cornell University

Before discussing the papers, I wish to make a few points. Statistics as a discipline should not be confused with statistics (numbers, adjectives) obtained from data. The distinction is not always made clear. We teach Statistics and not statistics. Statistics courses taught by academia are not necessarily appropriate for meeting the needs of users of statistical procedures. Many Statistics curricula do not teach how to obtain good national statistics. There is a fallacy that pervades the teaching of Statistics and that is that a good background in Mathematics is all one needs to teach Statistics courses. Teaching mathematical statistics and probability is completely different from teaching procedures and techniques for obtaining and handling large data sets. Most Statistics curricula concentrate on the former to the exclusion of the latter. Also, teaching students how to manipulate mathematical symbols is not teaching statistical theory.

In order to obtain good national statistics, it is necessary to establish criteria, goals, and rules of procedure which must be followed. The following items should be considered:

- a) Well defined objectives and populations -- Statisticians do not ordinarily precisely define a population for data from an investigation. For example, no textbook to my knowledge precisely defines the population structure for an experiment arranged in a ran-

domized complete block design. Statisticians do a lot of handwaving and dodge the issue by saying their inferences apply to the population (never defined) for which this investigation was a random sample. Such a procedure cannot be followed in obtaining reliable national statistics. The population and sampling procedure must be precisely defined.

- b) Accurate and reliable methods of measurement and quality control of the data -- If an investigator doesn't obtain good and reliable data, he is left with nothing but a bunch of numbers. No amount of statistical computation can add information to numbers. Too many investigations obtain numbers with little or no information in them. To illustrate, I was recently discussing a possible prenatal mortality study planned for Mexico and China. It was fairly obvious that the study as proposed would not result in much useful information. They would obtain a large set of numbers but rather poor data.
- c) Principles of scientific investigation -- These principles should be taught in research methods courses but if they are not it is up to the statistical community to do it. If these principles are not practiced, good scientific information will not result.
- d) Representative sampling procedures -- If the population is incompletely defined, how can one devise a sampling procedure to make the sample representative of the

population? Probability sample surveyors usually do a pretty fair job of defining their population but the remainder of statisticians do a lot of handwaving.

- e) Trained survey managers and interviewers -- To obtain good data and good national statistics, it is essential to have capable and reliable survey managers. They in turn will have well-trained and reliable interviewers. All aspects of a survey must be under control in order to obtain reliable data.
- f) Statistical computing procedures -- Accurate, reliable, and timely statistical computing procedures must be available in order to produce useful national statistics. Statistically competent personnel and procedures must be available. As an illustration, a nutrition study of fiber in the diet with 50 to 100 variables was conducted over ten years ago. Statistical analyses have not yet been developed to adequately analyze the data appropriately. Several nutrition theses have been written on the data using procedures not necessarily appropriate for the statistical design.
- g) Results in a usable and an intelligible form. In order for educators, politicians, administrators, and other consumers of statistics to understand the statistics presented, they must be in a comprehensible form. In order to convince lawmakers to continue funding a statistical organization, they must be convinced of its usefulness. This is an ongoing and continuous process.

If it is neglected, an organization may be dismantled because those in charge may believe that the funds could be put to better use elsewhere. To forestall this it is necessary to conduct an on-going campaign to keep administrators and politicians aware of the usefulness and need of the various national statistics.

Now let us turn to the particular papers. The paper by Kish et al. deals with recent developments in obtaining national statistics and to some extent with statistical offerings in universities in China. As pointed out by the authors, a lot is going on in China in obtaining national statistics. I would like to mention three additional investigations with which I am concerned. One is a cancer mortality dietary study on which T. Colin Campbell, Cornell University, and Junshi Chen, Beijing University, are Co-investigators. There are many statistical problems associated with this study. The second one is a selenium intervention study in China under the direction of L.C. Clark, Cornell University, and with the cooperation of a number of Chinese scientists. Several statistical problems are associated with this study. One mechanical problem bothering this study is how to mix selenium uniformly in salt. The salt crystals are not the same size and tend to separate during transportation. The other investigation is a hepatitis-liver cancer study on a large scale; L.C. Clark is involved here also. As pointed out by Kish et al. considerable progress has been made but much needs to be done.

The paper by Arnold compares statistical training at U.S. universities with training inside the country involved. He cites examples where U.S. trained statisticians tend to take jobs in a developed country rather than return to their native land. This points up the need for the government involved to provide a better environment for their statisticians and perhaps to do in-country training of their scientists. Perhaps their educational and statistical needs should be geared more to the needs of their country and not to the needs of the developed country where students obtain training. I would have liked to see more discussion on the type of training needed for statisticians in lesser developed countries. It is not clear that the standard Statistics curricula in the U.S. and other developed countries is at all adequate to meet the needs of lesser developed countries. Arnold believes that computing facilities are limiting in lesser developed countries. He does not say what facilities are required. For example, a good pocket calculator may be all that is required. We in the U.S. tend to think that what we use is what a lesser developed country needs. Commodore 64's around \$130 may be much more economical and useful than IBM PC's at approximately \$3,000.

I would have liked to hear more from Arnold on inadequacy of current statistical offerings for a developing country's needs. For example, the statistical design and analysis for intercropping experiments presents some real challenges and problems. These are illustrated to some extent in my forthcoming book on the subject. The question of quality of data is very important

in these countries.

The paper by McCracken was interesting in that the flavor was more of a user-of-statistics rather than a teacher-of-statistics one as were the first two papers. The training of interviewers, handling large data sets, etc. received much more attention in this paper than in the former two. Many of the points raised at the beginning of this discussion were considered by McCracken. The actual conduct of a survey needs much more attention in Statistics curricula than currently given. These curricula would be well-advised to incorporate training in areas mentioned by McCracken. Recently I presented a talk in which ten problems related to the future of Statistics were discussed. The first one was model formulation and selection and the second one was the training of survey statisticians. If statisticians do not meet the training needs of investigators in all areas, this training will be obtained elsewhere. If Statistics is to flourish, it should meet the needs of procurers and users of statistics. In this way we can be of help in obtaining good national statistics useful in making wise policy decisions.